

Verbetes

<http://services.sapo.pt>

@ InformationRetrievel/Verbetes

What?

What is Verbetes service?

Verbetes is a service that answers to requests of type "Who Is?" for public personalities.

What services are available?

Whols → Returns a list of jobs [Mourinho] [Mourinho @ 2010-01-01] [avançado]

GetPersonalities → Returns a list of names of public personalities [default] [min=10]

GetErgos → Returns a list of jobs [default] [XML]

How?

How does Verbetes works?

Verbetes is based on patterns extracted from news, applied to news on a hourly basis:

- “O primeiro ministro, José Sócrates, disse ontem ...”
- “Nuno Santos é o novo director de informação da RTP”

We extract alternative names like:

- “Alberto João Jardim” → “Jardim”
- “José Mourinho” → “Mourinho”
- “José Sócrates” → “pinócrates” (soon)

We also keep track of :

- the news title from where the verbete was extracted and its date/time
- the tags/labels given by the journalists

Numbers?

28,000 different verbetes

10,000 verbetes with 2 or more occurrences on news

1,200 verbetes with 5 or more occurrences on news

4,500 names with at least 1 alternative name

200 names with alternative names that occurs 5 or more times on news

Where?

Where can we use Verbetes?

@Voxx:

- help quotations extraction for VIP people
 - help quotations extraction where the name is hidden (“The president of BPN said that..”)
 - 30% of Voxx extracted quotations were based on Verbetes names and jobs
-
- To identify names of people on news or social networks
 - To identify names of organizations and companies
 - On the dataset of LUSA news (50 years of news)
 - Where there is a need of a historical point of view of a personality



Verbetes

<http://services.sapo.pt>
@ InformationRetrieval/Verbetes